U.S. DEPARTMENT OF COMMERCE
National Technical Information Service

PB-262 200

Numerical Optimization Techniques in Air Quality Modeling. Objective Interpolation Formulae for the Spatial Distribution of Pollutant Concentration

Carnegie-Mellon University, Pittsburgh, Pa.

Prepared for

Environmental Sciences Research Lab., Research Triangle Park, N.C.

Dec 76

021095

PB 262 200

EPA-600/4-76-058
December 1976

Environmental Monitoring Series

REPRODUCED BY
NATIONAL TECHNICAL
INFORMATION SERVICE
U. S. DEPARTMENT OF COMMERCE
SPRINGFIELD, VA. 22161

TECHNICAL REPORT DATA
(Please read Instructions on the reverse before completing)

1. REPORT NO.	2.	3. RECIPIENT'S ACCESSION NO.
EPA-600/4-76-058		

4. TITLE AND SUBTITLE	5. REPORT DATE
NUMERICAL OPTIMIZATION TECHNIQUES IN AIR QUALITY MODELING. Objective Interpolation Formulae for the Spatial Distribution of Pollutant Concentration	December 1976
	6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S)	8. PERFORMING ORGANIZATION REPORT NO.
S-A Gustafson, K. O. Kortanek and J. R. Sweigart	
9. PERFORMING ORGANIZATION NAME AND ADDRESS	10. PROGRAM ELEMENT NO.
Carnegie-Mellon University Pittsburgh, Pennsylvania 15213	1AA603 (1AA009)
	11. CONTRACT/GRANT NO.
	R 803632
12. SPONSORING AGENCY NAME AND ADDRESS	13. TYPE OF REPORT AND PERIOD COVERED
Environmental Sciences Research Laboratory Office of Research and Development U.S. Environmental Protection Agency Research Triangle Park, North Carolina 27711	Final 4/75-4/76
	14. SPONSORING AGENCY CODE
	EPA-ORD

15. SUPPLEMENTARY NOTES

16. ABSTRACT

A technique is proposed for objective interpolation of the air quality distribution over a region in terms of sparse measurement data. Empirical information provided by the latter is effectively combined with knowledge of atmospheric dispersion functions of the type commonly used in source-oriented air quality models, to provide improved estimates of the concentration distribution over an extended region. However, the technique is not primarily source-oriented since, in contrast to the real source distribution of a source-oriented model, it utilizes fictitious or pseudo-sources that are estimated in terms of the measured air quality data. This involves the use of interpolation functions that are computed using numerical optimization techniques based on the method of least squares. Due to the large number of different "weather" states that affect the atmospheric dispersion of pollution, considerable computation is required, although the bulk of this can be done in advance, so that the final interpolation from the measured values only requires very simple calculation. Thus the proposed method has the potential for application on a real-time basis.

In addition to the mathematical formulation of the problem, this preliminary study includes some numerical experiments, using a current multiple-source EPA air quality model, to illustrate the technique.

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group
* Air pollution * Atmospheric composition Meteorological data * Numerical analysis * Interpolation * Least squares method * Empirical equations * Mathematical models		13B 04A 04B 12A

18. DISTRIBUTION STATEMENT	19. SECURITY CLASS *(This Report)*	21. NO. OF PAGES
RELEASE TO PUBLIC	UNCLASSIFIED	51
	20. SECURITY CLASS *(This page)* UNCLASSIFIED	22. PRICE

EPA Form 2220-1 (9-73)

RESEARCH REPORTING SERIES

Research reports of the Office of Research and Development, U.S. Environmental Protection Agency, have been grouped into five series. These five broad categories were established to facilitate further development and application of environmental technology. Elimination of traditional grouping was consciously planned to foster technology transfer and a maximum interface in related fields. The five series are:

1. Environmental Health Effects Research
2. Environmental Protection Technology
3. Ecological Research
4. Environmental Monitoring
5. Socioeconomic Environmental Studies

This report has been assigned to the ENVIRONMENTAL MONITORING series. This series describes research conducted to develop new or improved methods and instrumentation for the identification and quantification of environmental pollutants at the lowest conceivably significant concentrations. It also includes studies to determine the ambient concentrations of pollutants in the environment and/or the variance of pollutants as a function of time or meteorological factors.

This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161.

EPA-600/4-76-058
December 1976

NUMERICAL OPTIMIZATION TECHNIQUES IN AIR QUALITY MODELING
Objective Interpolation Formulae for the Spatial Distribution of Pollutant Concentration

by

S-Å Gustafson
The Royal Institute of Technology
Stockholm

K. O. Kortanek
Carnegie-Mellon University
Pittsburgh, Pennsylvania 15213

J. R. Sweigart
University of South Carolina
Columbia, South Carolina 29208

Grant No. R 803632

Project Officer

Kenneth L. Calder
Meteorology and Assessment Division
Environmental Sciences Research Laboratory
Research Triangle Park, North Carolina 27711

ENVIRONMENTAL SCIENCES RESEARCH LABORATORY
OFFICE OF RESEARCH AND DEVELOPMENT
U.S. ENVIRONMENTAL PROTECTION AGENCY
RESEARCH TRIANGLE PARK, NORTH CAROLINA 27711

DISCLAIMER

This report has been reviewed by the Environmental Sciences Research Laboratory, U.S. Environmental Protection Agency, and approved for publication. Approval does not signify that the contents necessarily reflect the views and policies of the U.S. Environmental Protection Agency, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

ABSTRACT

A technique is proposed for objective interpolation of the air quality distribution over a region in terms of sparse measurement data. Empirical information provided by the latter is effectively combined with knowledge of atmospheric dispersion functions of the type commonly used in source-oriented air quality models, to provide improved estimates of the concentration distribution over an extended region. However, the technique is not primarily source-oriented since, in contrast to the real source distribution of a source-oriented model, it utilizes fictitious or pseudo-sources that are estimated in terms of the measured air quality data. This involves the use of interpolation functions that are computed using numerical optimization techniques based on the method of least squares. Due to the large number of different "weather" states that affect the atmospheric dispersion of pollution, considerable computation is required, although the bulk of this can be done in advance, so that the final interpolation from the measured values only requires very simple calculation. Thus the proposed method has the potential for application on a real-time basis.

In addition to the mathematical formulation of the problem, this preliminary study includes some numerical experiments, using a current multiple-source EPA air quality model, to illustrate the technique that is proposed.

CONTENTS

Abstract iii

Acknowledgement vi

1. Introduction 1

2. Multiple-Source Plume Models and the Concept of Pseudo Sources 4

3. Further Development of the Pseudo-Source Concept . . 8

4. The Interpolation Formula for Concentration 13

5. Interpolation Formula for Multiple Weather States . . 22

6. Least Squares Computation for Interpolants 26

7. A Numerical Experiment in Concentration Interpolation 30

8. The Same Example with Aggregation of Pseudo-Sources . 37

9. Recommendations 41

References 42

Preceding page blank

ACKNOWLEDGEMENTS

The authors gratefully acknowledge several extended discussions with the Project Officer, which led to identification of the air quality interpolation problem as one particularly appropriate for analysis by numerical optimization techniques. Mr. K. L. Calder also had a major part in clarifying some of the interpretations and developing the final report on the research, based on an early draft.

The authors are also grateful to Professor E. L. Rubin, Carnegie-Mellon University for communications regarding a hypothetical 25-point source inventory used in numerical experiments. J. R. Sweigart received financial support during the year as a Richard King Mellon Fellow in Environmental Studies, Carnegie-Mellon University. Finally, we are indebted to R. Edahl and D. N. Lee (also a Richard King Mellon Fellow) for their assistance in the computational experiments.

SECTION 1

INTRODUCTION

Even when attention can be restricted to near-ground values of the pollutant concentration, the distribution of air quality over an urban area represents a two-dimensional time-dependent scalar field of great complexity. Any proper understanding of its complex structure must involve theoretical considerations, like those in mathematical source-oriented air quality models which estimate the air quality in terms of specified pollutant emissions and appropriate atmospheric dispersion functions, or alternatively the use of measured air quality data based on the operation of an extensive sampling network. If a mathematical air quality model is available, it will generally be possible to estimate the pollutant concentration at an arbitrary receptor location, and hence with sufficient computational effort to determine the air quality distribution over the entire area. However, in contrast, when only empirical air quality data from a sampling network of fixed stations is available, it will normally be necessary to interpolate concentration values for the intermediate locations, and in this case it is desirable that an objective procedure be used, i.e., one that does not involve personal or subjective judgment.

The preliminary analysis of the present paper shows that

these two approaches can be complementary and that real air quality measurements at typical locations can be combined with appropriate atmospheric dispersion functions to provide an air quality interpolation function for an entire region. The latter does not require complete information on the real source strengths of the pollutant emissions such as would be required for application of a conventional source-oriented air quality model. The interpolation function, however, actually involves the concept of a hypothetical or fictitious "pseudo-source" distribution, that is computed for a given atmospheric dispersion situation or weather state, by optimized fitting techniques from the measured air quality values at selected stations and the specified atmospheric dispersion functions. Similar ideas have recently been exploited by Heimbach and Sasaki (1975) in fitting an analytical dispersion model to sparse data on air quality. Their work was motivated by the fact that detailed emission inventories are frequently difficult to obtain, and the predictions of source-oriented air quality models should be capable of improvement when air quality measurements are available.

Since the interpolation approach is self-correcting, the air quality estimates are less affected by inaccuracies in the real emissions data. In a sense, it provides a technique for adjusting or "calibrating" the air quality estimates provided by a conventional source-oriented model in order to correct for errors in the estimates of the emissions. The calibration technique is, however, not subject to the fundamental objection

that can be raised in connection with some others that have been used with source-oriented models (Brier, 1973). It is, in principle, the complement of another approach using empirical data recently suggested by Calder et.al. (1975). In the latter, multiple-station observations of air quality are coupled with estimates of the emissions for a multiple-source distribution, to deduce the effective atmospheric dispersion functions. The latter can then be used as a basis for a conventional source-oriented air quality model. The pseudo-source interpolation technique, on the other hand, estimates "effective" emissions in terms of the observed air quality and prescribed atmospheric dispersion functions. In principle, the end result is a "calibrated" interpolatory air quality model. Since the interpolation functions can be computed in advance, the technique has potential for use in air quality prediction on a real-time basis.

SECTION 2

MULTIPLE-SOURCE PLUME MODELS AND THE CONCEPT OF PSEUDO-SOURCES

The general structure of multiple-source plume models for urban air pollution has recently been discussed by Calder (1976). The starting point for most models is the working assumption that the long-term variability of emissions, air quality and atmospheric conditions can be treated as though it resulted from a time-sequence of different steady-state situations. The sequence interval is normally taken to be relatively short, and perhaps only of the order of an hour. For pollutants that can be regarded as chemically inert, it is further assumed that the concentration contributions produced at any receptor location from several sources combine additively, so that the individual plumes can be simply superimposed. Then the total concentration $C(x)$ at the point x can be written

$$C(x) = \sum_{j=1}^{N} q_j v_j(x), \qquad (1)$$

where

q_j is the strength of source j

v_j is the concentration produced by a <u>unit source</u> at the location of j.

The function v_j is, of course, independent of source strength and is assumed to be a computable function of the meteorological

conditions, such as wind direction, wind speed, atmospheric stability, mixing depth, etc. The particular combination of these conditions existing during any one steady-state interval of the time sequence is regarded as constituting one of an ensemble of possible meteorological or "weather" states. The expression of the total concentration $C(x)$ as a linear superposition extending over the atmospheric dispersion functions $v_j(x)$ is mathematically analogous to the use of polynominal or trigonometric functions for finite expansion approximation of continuous functions.

Before considering the general pseudo-source concept, we first consider a very simple example to illustrate the basic concept and its use for interpolating air quality over a region in terms of a limited number of observed values. We consider a single source of strength q at $x = x_o$, and a hypothetical one-dimensional concentration distribution $C(x)$ produced by it. (Note that here x is not to be confused with that in equation (1) above, where it denotes a general position vector). Then, corresponding to just a single term of the sum in (1), we have

$$C(x) = qv(x-x_o) \tag{2}$$

so that $C(x)$ can be determined once q is specified. However, in the absence of direct knowledge of q, it may be estimated in terms of a "measured" concentration value, say C_1 at location $x = x_1$, and the corresponding value $v(x_1-x_o)$ of the dispersion functions, since $q = C_1/v(x_1-x_o)$. This value constitutes a

"pseudo-source" strength which when substituted into the equation for $C(x)$ then provides the general interpolation formula

$$\hat{C}(x) = C_1 v(x-x_o)/v(x_1-x_o). \tag{3}$$

This might be written $\hat{C}(x) = m(x)C_1$, where $m(x) = v(x-x_o)/v(x_1-x_o)$ is a simple interpolation function that "converts" the measured value C_1 into a field of values corresponding to the different values of x. The function $m(x)$ is completely independent of the actual strength of the source and depends only on the positional parameters of the source and the meteorological conditions that affect the dispersion function. Thus the interpolation function could be calculated once and for all without regard to the value of the pseudo-source strength $C_1/v(x_1-x_o)$.

This simple idea can be immediately generalized for equation (1). Thus if concentrations are measured at p sampling stations $x_1, x_2, \ldots, x_p$ yielding values C_i, $i = 1,2,\ldots,p$, then we have the following linear system to determine the N pseudo-source strengths q_j $(j = 1,2,\ldots,N)$

$$\sum_{j=1}^{N} q_j v_j(x_i) = C_i, \quad i = 1,2,\ldots,p. \tag{4}$$

Under appropriate conditions this system of equations will determine a set of pseudo-source values, which we will denote $\hat{q}_j$. On

substituting these back into equation (1) we obtain the desired air quality interpolation formula for the region, namely

$$\hat{C}(x) = \sum_{j=1}^{N} \hat{q}_j v_j(x). \qquad (5)$$

The "hat" notation is used here, since as will be seen later, it will sometimes be necessary to consider inconsistent sets of equations having no exact mathematical solutions, but only approximate solutions as, for example, in a least squares sense. In this case, of course, the values of $\hat{C}(x_i)$ for the sampling locations x_i will not exactly duplicate the measured values C_i, $i = 1,2,\ldots,p$. Finally, we note that in principle the dispersion functions $v_j(x)$ could be precalculated, once and for all, for a grid of points. The interpolated concentration distribution could then be very simply calculated from (5) in terms of the q_j, which from (4) are functions only of the observed values C_i and $v_j(x_i)$ $(i = 1,2,\ldots,p)$. This offers the hope for interpolation on a real-time basis. The remainder of the paper in effect only develops these simple concepts further. The complications arise from the need to consider many sources and the use of approximate solutions in a least-squares sense, together with the need for consideration of an ensemble of possible weather states.

SECTION 3

FURTHER DEVELOPMENT OF THE PSEUDO-SOURCE CONCEPT

The number of real identified sources, N in equation (1) may be very large for an urban area. However, an aggregated pseudo-source formulation in terms of a much smaller number of sources $n < N$ can be developed. Formally, we proceed as follows: Let $j_0, j_1, \ldots, j_n$ be integers such that

$$0 = j_0 < j_1 < \ldots < j_n = N.$$

Then we combine sources with indices between $j_0 + 1$ (= 1) and j_1 into the first pseudo-source, those with indices from $j_1 + 1$ to j_2 into the second pseudo-source, and so on. Generally, pseudo-source number r consists of sources with indices from $j_{r-1} + 1$ to j_r. Let the concentration contribution from the r-th pseudo-source be U_r. Thus (1) becomes

$$C(x) = \sum_{r=1}^{n} U_r(x), \qquad (6)$$

where

$$U_r(x) = \sum_{j=j_{r-1}+1}^{j_r} q_j v_j(x).$$

We now write $U_r(x)$ in the following manner:

$$U_r(x) = Q_r u_r(x),$$

where

$$Q_r = \sum_{j_{r-1}+1}^{j_r} q_j, \quad u_r(x) = U_r(x)/Q_r. \tag{7}$$

Q_r is the combined strength of all the sources in source class r. The functions $u_r(x)$ are still dependent on source-strength considerations since they depend on the ratio of the individual source strengths in the class, to the class total. This is in contrast to the $v_j(x)$ functions of equation (1) which are <u>true</u> dispersion functions that are entirely independent of source strength considerations. The pseudo-source concept is only entirely independent of knowledge of the real source distribution in the obvious case $n = N$, $j_r = r$, when $Q_r = q_r$, so that $u_r(x) = v_r(x)$. However the determination of the large number of values of q_j $(j = 1,2,\ldots,N)$ in (5) by the pseudo-source method could be computationally prohibitive, and this is the primary reason for the aggregated source concept.

Another case is very important as an approximation. For it is obvious that the real sources may be aggregated into classes in such a manner that within each class the strengths are "approximately" equal, i.e., to within some prescribed variation. This motivates an important specialization of the aggregated source concept, for which every source in a class is replaced by the average value, say $\bar{q}(r)$, for the class. This approximation

can be made as close as we please up to the limit of the completely unaggregated situation. It will be designated as the aggregated-averaged approximation (A.A.A.). For this case we see from (6) that

$$U_r(x) \doteq \bar{q}(r) V_r(x); \; V_r(x) = \sum_{r \text{ class}} v_j(x), \tag{8}$$

and the function $V_r(x)$ is now completely independent of source-strength considerations. Like the original functions $v_j(x)$ it depends only on positional parameters for the sources constituting the class, and on the meteorological conditions as affecting dispersion. Evidently from (7) the same will be true for the functions

$$u_r(x) = V_r(x)/\text{No. in r-class.}$$

In terms of the aggregated pseudo-source classes equation (1) is now replaced by a sum that only involves n terms, namely

$$C(x) = \sum_{r=1}^{n} Q_r u_r(x). \tag{9}$$

Now, if concentrations are measured at p sampling stations $x_1, x_2, \ldots, x_p$ yielding values C_i, $i = 1,\ldots,p$, we have the following linear system of equations to determine the n pseudo-source strengths Q_r $(r = 1,2,\ldots,n)$

$$\sum_{r=1}^{n} Q_r u_r(x_i) = C_i, \quad i = 1,2,\ldots,p. \tag{10}$$

If $p < n$, so that there are fewer equations than unknowns, then the system is underdetermined and will not suffice for a unique determination of the n unknown pseudo-source strengths Q_r $(r = 1,2,\ldots,n)$. If $p = n$, so that there are as many equations as unknowns, the system is even determined and, except in a singular case (i.e., vanishing of the determinant of the coefficients), will determine the Q_r's uniquely. If $p > n$, so that there are more equations than unknowns, the system is overdetermined and there will generally be no exact mathematical solution for the Q_r's. However, in the latter case, there will still be a best approximate solution in a least squares sense (cf., Lanczos Applied Analysis, Prentice-Hall, 1965) for which

$$\Delta = \min_{\vec{Q}} \sum_{i=1}^{p} \Big[\sum_{r=1}^{n} Q_r u_r(x_i) - c_i \Big]^2, \tag{11}$$

and the vector $\vec{Q} = (Q_1, Q_2, \ldots, Q_n)$ is such as to minimize Δ. Thus in the even determined case we have a mathematically exact solution for the Q_r's and for the overdetermined case a least squares approximation. If $\hat{Q}_r$ denotes the appropriate solution for these two cases, then on substituting back in (9), we obtain the corresponding air quality interpolation formula for the entire concentration field based on the use of pseudo-sources, viz.,

$$\hat{c}(x) = \sum_{r=1}^{n} \hat{Q}_r u_r(x). \tag{12}$$

This is a solution in the least-squares sense of equation (11),

namely

$$\sum_{i=1}^{p} [\hat{c}(x_i) - c_i]^2 = \text{minimum.}$$

It should be noted that equation (12) is analogous to equation (5) for the example that was considered earlier. The case of a uniform background concentration B extending over the whole region is realized by setting $Q_1 = B$ and $u_1(x) = 1$ for all x.

SECTION 4

THE INTERPOLATION FORMULA FOR CONCENTRATION

Still restricting attention to the interpolation problem for a single meteorological or weather state, we next consider the form of the interpolation formula, for the two cases just considered of an even-determined and over-determined system. The discussion in this Section is based on well-known results in linear algebra. See, e.g., Dahlquist et.al., 1974, Chapter 5.

EVEN-DETERMINED SYSTEM

We consider the set of equations (10) for the case $p = n$. This set has a unique solution for the Q_r's provided the determinant of its matrix of coefficients is not zero (a matrix having determinant zero corresponds to the singular case). We treat $u_r(x_i)$ $[r = 1,2,\ldots,n;\ i = 1,2,\ldots,p]$ as a $(n\times p)$ matrix, say U_{ri}. Let $U_{ri} = U^T_{ir}$ = say A_{ir} (so that $A = U^T$ is a $(p\times n)$ matrix). Then (10) becomes

$$C_i = \sum_{r=1}^{n} A_{ir}Q_r, \quad i = 1,2,\ldots,p, \tag{13}$$

or in matrix notation

$$C = AQ, \tag{14}$$

where generally C and Q are column vectors of p and n

elements respectively. We solve this equation for Q (say $\hat{Q}$) and then set

$$\hat{c}(x) = \sum_{r=1}^{n} \hat{Q}_r u_r(x) = u^T(x)\hat{Q} \qquad (15)$$

to obtain the required interpolation formula for the entire concentration field $\hat{c}(x)$. [For given x, $u(x)$ is a column vector of n elements, so $u^T(x)$ is a row vector of n elements, while Q is a column vector of n elements. For given x, $\hat{c}(x)$ is a one-element matrix or scalar, in contrast to the vector C.] Equation (15) is, of course, equivalent to (12). We now show that (15) can be written in a different, although exactly equivalent form as follows:

$$\hat{c}(x) = u^T(x)\hat{Q} = u^T(x)A^{-1}C = [C^T(A^{-1})^T u(x)]^T$$

since for 3 matrices L, M, N, $(LMN)^T = N^T M^T L^T$ where the superscript T denotes a transpose. Hence

$$\hat{c}(x) = [C^T m(x)]^T = m^T(x)C, \qquad (16)$$

where

$$m(x) = (A^{-1})^T u(x). \qquad (17)$$

Here for given x, $m(x)$ is a column vector of $p(= n)$ elements $m_i(x)$. Put $B = (A^{-1})^T = (A^T)^{-1}$ so that $m(x) = Bu(x)$. Hence

$$m_i(x) = \sum_{r=1}^{n} B_{ir} u_r(x), \quad i = 1,2,\ldots,p.$$

Also, since $A^T m(x) = u(x)$,

$$u_r(x) = [A^T m(x)]_r = \sum_{i=1}^{p} A^T_{ri} m_i(x) = \sum_{i=1}^{p} A_{ir} m_i(x)$$

$$= \sum_{i=1}^{p} m_i(x) u_r(x_i), \quad r = 1,2,\ldots,n. \tag{18}$$

Equation (16) in expanded form is

$$\hat{C}(x) = \sum_{i=1}^{p} m_i(x) C_i, \tag{19}$$

which is thus the required interpolation formula (15). Since for the even-determined case under consideration the solution of (14) for $\hat{Q}$ is exact, in this case we have $C(x_i) = C_i$ ($i = 1,2,\ldots,p$), at the sampling locations where the C_i are the measured values. We note that equation (18) can be computed without any direct determination of the pseudo-source strengths.

Two very important features should be noted from the preceding analysis:

(a) Since the functions $u_r(x)$ become "true" dispersion functions when the sources are unaggregated (i.e., $n = N$), and also when the "aggregated-average approximation" is used, the interpolation functions $m_i(x)$ satisfying the system (18) will be completely independent of source-strength considerations under these conditions, and will only depend on the source positional parameters and meteorological factors. Thus the interpolation formula should be valid for arbitrary

values of the source strengths at the specified locations of the real source distribution. Likewise, the concentration interpolation formula (19) should be valid for arbitrary values of the source strengths.

(b) From equation (18) it follows that the $m_i(x)$ do not depend on the source class r, i.e., they serve as general interpolation functions for all the $u_r(x)$ functions. Since the individual functions $u_r(x)$ may be widely different for different values of r, it follows that the interpolation functions $m_i(x)$ must be insensitive to the form of the functions $u_r(x)$. In this case it may be conjectured that a similar simple relation might also be applicable even when several different weather states are involved, as when these are appropriately aggregated into a single weather class. As we shall see later, the aggregation of different weather states into weather classes for which an equation like (18) for the interpolation $m_i(x)$ may be an acceptable approximation must be the subject of empirical study, although this is the basis for analysis of the multiple-weather state interpolation problem.

OVER-DETERMINED SYSTEM

For $p > n$, the system becomes over-determined and in this case we seek for a solution Q that is the best solution in a least squares sense. Then an argument similar to that in the preceding section may be used. Thus as before we take

$$\hat{C}(x) = u^T(x)\hat{Q},$$

where $\hat{Q}$ is the best solution in the least squares sense of

$$C = AQ,$$

where the matrix A has p rows and n columns. This least squares solution is the solution of (cf. Lanczos, loc cit)

$$A^TAQ = A^TC.$$

The remarkable fact about this equation is that it gives an <u>even-determined</u> system having just as many equations as unknowns, no matter how strongly over-determined was the original system. Its solution is*

$$\hat{Q} = (A^TA)^{-1}A^TC. \tag{20}$$

Then

$$\begin{aligned}\hat{C}(x) = u^T(x)\hat{Q} &= u^T(x)(A^TA)^{-1}A^TC \\ &= [C^TA(A^TA)^{-1}u(x)]^T \\ &= [C^Tm(x)]^T = m^T(x)C,\end{aligned} \tag{21}$$

where

$$m(x) = A(A^TA)^{-1}u(x) \tag{22}$$

---

*Note that now $A\hat{Q} = A(A^TA)^{-1}A^TC \neq C$.

and, for given x, is a column vector with p elements $(m_i,\ i = 1,2,\ldots,p)$. Multiplying by A^T we obtain

$$A^T m(x) = u(x). \tag{23}$$

The system (23) always has the exact solution given by (22), but is generally not unique since (23) has p unknowns $(m_i,\ i = 1,2,\ldots,p)$, but only n equations. However, if $m(x)$ is a solution of (23), then we set

$$\hat{C}(x) = m^T(x)\,C = \sum_{i=1}^{p} m_i(x)\,C_i \tag{24}$$

as before in equation (19). A particular solution obtained from least squares is, of course, (22) and gives the least squares estimate $\hat{C}(x)$. From the above analysis, it is seen that even in the over-determined case, provided a least squares solution is adopted, then the same form of concentration interpolation formula is available. However, this is now an approximation in a least squares sense, so that $C(x_i) \neq C_i$.

Finally, we may note the underdetermined case mentioned in Section 3, where for equation (10) (or equivalently (14)) $p < n$, so that in general the system has many solutions. This implies that the <u>adjoint</u> system to (14), namely, the system (23) is over-determined and has no exact solution. It then follows from the preceding analysis for the over-determined case, that the least-squares solution of (23) (compare with (20)) is

$$m(x) = (AA^T)^{-1}Au(x)$$

under the assumption that the rank of the (p×n) matrix A is p. A corresponding exact but non-unique solution for the unknown pseudo-source strengths $\hat{Q}_r$ from the underdetermined system (14) is then

$$\hat{Q} = A^T(AA^T)^{-1}C.$$

Now as before on substituting for $\hat{Q}$, the formula $\hat{e}(x) = u^T(x)\hat{Q}$ becomes

$$\hat{e}(x) = u^T(x)\hat{Q} = u^T(x)A^T(AA^T)^{-1}C = m^T(x)C,$$

i.e.,

$$\hat{e}(x) = \sum_{i=1}^{p} m_i(x)C_i.$$

Thus the underdetermined case can be analyzed by the same least-squares procedure. A classification of the solutions of the systems (14) and (23) in terms of the above discussion can be summarized as below.

Case	Rank Assumption for A	Equation (14)	Equation (23)
		$AQ = C$ or $\sum_{r=1}^{n} Q_r u_r(x_i) = C_i$ $A = (p \times n)$; $Q = (n \times 1)$ $C = (p \times 1)$	$A^T m = u$ or $\sum_{i=1}^{p} u_r(x_i) m_i(x) = u_r(x)$ $A^T = (n \times p)$; $m = (p \times 1)$ $u = (n \times 1)$
(1) Overdetermined $p > n$	n	No solution (exact) Define (least-squares) $\hat{Q} = (A^T A)^{-1} A^T C$	Has solutions Choose $m(x) = A(A^T A)^{-1} u(x)$
(2) Underdetermined $p < n$	p	Has solutions Choose $\hat{Q} = A^T (AA^T)^{-1} C$	No solution (exact) Define (least-squares) $m(x) = (AA^T)^{-1} Au(x)$

As for the even-determined situation previously considered, the interpolation function $m_i(x)$ (as given by equation (23), which in expanded form is equation (18)) should be entirely independent of source-strength considerations when there are either unaggregated or aggregated-averaged pseudo sources. Similarly, the concentration interpolation formula for $C(x)$ should be valid independent of source-strength considerations. However, as we are now dealing with equations that are only satisfied in a least-squares sense of approximation, it is necessary to check empirically to determine exactly how good such "approximate" conclusions may be.

SECTION 5

INTERPOLATION FORMULA FOR MULTIPLE WEATHER STATES

In the preceding discussion we have been concerned with a single (steady-state) meteorological dispersion condition, that involves only a single set of source-receptor functions $u_r(x)$, $r = 1,\ldots,n$. Normally, however, multiple-source air quality modeling will involve consideration of many different meteorological conditions or "weather states" (Calder, 1976).

Let Ω be the set of all weather states w, each say as specified by wind direction, wind speed, stability class and mixing depth. For example, one might specify 16 wind directions, 10 wind speed classes, 6 stability classes and 4 different mixing depths (cf. Calder, loc.cit.). Then Ω contains $16 \cdot 10 \cdot 6 \cdot 4 = 3840$ weather states. For each of these we shall have an equation of the form (9), namely

$$C^w(x) = \sum_{r=1}^{n} Q_r u_r^w(x). \qquad (25)$$

In principle we could now proceed exactly as before and for each weather state determine the appropriate interpolation function $m(x)$ as a basis for the general interpolation formula $\hat{C}^w(x)$ for the concentration at an arbitrary location x for the weather state w occurring. However, in view of the large number of weather states this could involve a prohibitive amount of

computation. We thus consider the conjecture mentioned in Section 4, and based upon the observation that for a given weather state the single interpolation function $m(x)$ applies for all the (possibly widely) different source-receptor functions $u_r^w(x)$ in a class r. We therefore consider a partitioning of the set of all weather states Ω into a smaller number of subsets or weather classes

$$\Omega_1, \Omega_2, \ldots, \Omega_j,$$

and examine the question of whether an adequate general concentration interpolation formula can be derived for $\bar{C}^w(x)$ in terms of interpolation functions $m(x)$ calculated for each of the smaller number of classes, rather than the large number of individual states.

As partitioning is a combinatorial problem it may be accomplished in many ways, and a fuller study would be required to compare the various possibilities. Only a very limited initial study is presented in this report. However, for each weather class Ω_k of a partitioning we shall need to examine whether a single interpolation function $m(x)$ can be employed (to avoid confusion we refrain from using the index symbol k on $m(x)$) such that (see equations (23) and (24))

$$A^T m(x) = u^w(x)$$

or

$$\sum_{i=1}^{p} u_r^w(x_i)\, m_i(x) = u_r^w(x); \quad r = 1,2,\ldots,n \qquad (26)$$

all w in Ω_k,

and

$$\hat{C}^w(x) = m^T(x)\, C^w = \sum_{i=1}^{p} m_i(x)\, C_i^w, \text{ all } w \text{ in } \Omega_k. \qquad (27)$$

Here C_i^w denotes the measured concentration value at the sampling station i ($i = 1,2,\ldots,p$) for the weather state w that is occurring. It may be noted that (27) follows immediately from (26) on using the pseudo-source strengths $\hat{Q}_r$ defined by the solution of (25).* For

$$\hat{C}^w(x) = \sum_{r=1}^{n} \hat{Q}_r u_r^w(x) = \sum_{r=1}^{n} \hat{Q}_r \Big\{ \sum_{i=1}^{p} u_r^w(x_i)\, m_i(x) \Big\}$$

$$= \sum_{i=1}^{p} m_i(x) \Big\{ \sum_{r=1}^{n} \hat{Q}_r u_r^w(x_i) \Big\}$$

$$= \sum_{i=1}^{p} m_i(x)\, C_i^w.$$

The basic concept of the pseudo-source technique for the multiple-weather state situation is thus to determine the interpolation functions $m_i(x)$ by least-squares solution (if necessary) of

---

*To reduce to the very simple example of Section 2, we take $n = 1 = p$, so $u_r^w(x) \equiv v(x-x_o)$ and $m(x) \equiv v(x-x_o)/v(x_1-x_o)$. Then (26) corresponds to $v(x-x_o) = m(x)\, v(x_1-x_o)$ and (27) to $\hat{C}(x) = m(x)\, C_1$.

the system (26), and then substitute into (27) to obtain the required interpolation formula. As we have noted the method can be applied irrespective of whether the system is over-, even- or underdetermined. In a fuller study it would be very desirable to examine in detail the adequacy of the tentative conclusion that the interpolation functions should be entirely independent of source-strength considerations and only dependent on the source positional parameters and the weather class. Because of the approximations that have been noted this conclusion may only be valid in an approximate sense, and the degree of approximation must be tested experimentally.

SECTION 6

LEAST SQUARES COMPUTATION OF INTERPOLANTS

Calculation of the interpolation functions $m_i(x)$ from (26) will first require calculation of the functions $u_r^w(x)$ [$r = 1,2,\ldots,n$, all w in Ω_k]. As was shown in Section 3, these functions reduce to "true" dispersion functions when the sources are unaggregated, i.e., $n = N$, so that

$$u_r^w(x) = v_r^w(x), \quad r = 1,2,\ldots,N,$$

and in the special case of the aggregated-averaged approximation, when

$$u_r^w(x) = \sum_{r \text{ class}} v_j^w(x)/\text{No. in class}, \quad r = 1,2,\ldots,n(< N).$$

The least squares solution of (26) will then require the consideration of a set of $n \times s_k$ equations for the p unknown interpolants $m_i(x)$, where s_k denotes the number of weather states in the weather class Ω_k. In contrast, the system (27), which as we have just seen is actually the direct result of summing (26) over all the n source classes after first weighting by the corresponding pseudo-source strengths Q_r, is a system of only s_k equations for the interpolants, which will therefore be much simpler for least-squares computation. In the present exploratory study, as discussed in more detail in Section 7,

concentration estimates were generated using an existing EPA source-oriented plume model of the form

$$C^w(x) = \sum_{j=1}^{N} q_j v_j^w(x) \tag{28}$$

with an arbitrary source-strength distribution q_j ($j = 1,2,\ldots,N$). For this reason also it was convenient to use the direct output of the model to determine the interpolants $m_i(x)$ by least squares applied directly to equations (27) and using the model to provide estimates of both $\hat{C}^w(x)$ and C_i^w. As we have seen, in some approximate sense that has yet to be studied in detail, the interpolants should be independent of the assumed source distribution and thus generally applicable to any distribution of source strengths at the specified locations.

Before considering the numerical experiments, we first consider some general aspects of the least-squares problem for the system (27) and where the $C^w(x)$ values are provided by a model as in (28). We first select a finite set of points x, like the points of a rectangular grid R, at which we eventually desire to interpolate values of the concentration $\hat{C}^w(x)$ for any weather state w, using only the values C_i, $i = 1,2,\ldots,p$ at the p sampling stations x_i. The number of weather states in the weather class Ω_k is s_k. Thus for any fixed point x of R we consider the set of linear equations

$$\sum_{i=1}^{p} z_i C^w(x_i) = C^w(x), \quad w \in \Omega_k. \tag{29}$$

This is a set of s_k equations for the p unknowns z_i, $i = 1,2,\ldots,p$ and we will consider the over-determined case $s_k > p$. Such equations may be solved in a least squares sense. The solution z_i defines a point in the interpolation function table $m_i(x)$ (where the dependence of the functions on the weather class Ω_k is not explicit in the notation).

The amount of computational effort is illustrated by returning to the Ω above, containing 3840 weather states. We let Ω_k contain all states with a fixed wind direction. Then Ω is split into 16 subgroups, each containing 240 states. Assume now that we have 10 sampling stations and that the interpolation grid contains 400 points. Hence for each point x of the grid we must solve 16 over-determined systems of equations (29) with 10 variables z_i and 240 equations. That is, for each x of R we need to store $10 \times 16 = 160$ coefficients $m_i(x)$, or a total of 64,000 numbers for the entire grid. This computational effort is formidable although it only needs to be done once. Then when concentration values are available from actual measurements C_i at a limited number of sampling stations during weather state w, the concentration field over the entire grid may be very simply calculated in terms of just these p sampled values from (27). The important feature of the method is that it would permit interpolation on a real-time basis, since the functions $m_i(x)$ could be computed in advance.

Finally, we may note in the above example that if there had been $N = 200$ unaggregated sources in the region and the system

(26) has been used for the least-squares determination of the interpolants, then this would still have involved the 10 variables z_i, but now $240 \times 200 = 48{,}000$ equations. If the 200 real sources had been aggregated into 20 pseudo-source classes (so $n = 20$) with dispersion functions $u_r^w(x)$ corresponding to the aggregated-averaged approximation as above, then only $240 \times 20 = 4800$ equations would be involved for the 240 weather states in the weather class. The storage requirements for the interpolants $m_i(x)$ are, of course, the same as previously in both cases, i.e., for each x of R we need 160 coefficients, or a total of 64,000 for the entire grid for the one weather class Ω_k.

In the least-squares computations performed in Sections 7 and 8, a linear regression package, IAREG, of the Graduate School of Industrial Administration Program Library at Carnegie-Mellon University was used. These statistical programs are available both for the IBM 360/Time Sharing System and the UNIVAC 1108 at the Carnegie-Mellon Computation Center. IAREG is a FORTRAN program developed in 1974 by Bouwman and Prescott at Carnegie-Mellon. In addition to providing the standard goodness of fit measures such as the means and standard deviations of the estimated coefficients, it prints the correlation matrix of the estimated coefficients and also plots the normal probabilities of the residuals. The package also performs adaptive regression and multiple-regression.

SECTION 7

A NUMERICAL EXPERIMENT IN CONCENTRATION INTERPOLATION

In lieu of real measurement data a preliminary numerical experiment was performed using synthetic or simulated air quality data generated by a current EPA multiple-source model (EPA code-DBT51). An interactive version of this is available on the EPA UNAMAP (Users' Network for Applied Modeling of Air Pollution) computer system. This model is a point-source model that produces hourly concentration values for up to 30 receptors whose locations are specified, and for up to 25 sources. It is based on a Gaussian plume dispersion formulation. Inputs to the program consist of the number of sources to be considered, and for each the emission rate, physical height of emission, stack gas temperature, volume flow or stack gas velocity and diameter, and the coordinates of the source location. The numbers of receptors, their coordinates and heights above ground level, are also required. For each hour the meteorological or weather state is specified in terms of wind direction, wind speed, stability class, mixing height (and ambient air temperature which is required in the estimation of plume rise).

The above multiple-source code was applied to generate synthetic concentration data for a hypothetical air quality region containing 25 major point sources, and under the assumption of

zero background concentration. Although hypothetical, the source inventory reflects a realistic spatial distribution of sources and stack characteristics. Likewise, the locations of 3 hypothetical measurement stations were selected in an ad hoc fashion. The source inventory is given in Table 1, and the DBT51 code then provided synthetic concentration estimates corresponding to this inventory by the superposition of 25 "Gaussian" plumes in the form of equation (28), namely

$$c^w(x) = \sum_{j=1}^{25} q_j v_j^w(x).$$

Figure 1 shows:

(a) the approximate locations of the individual sources on a 15 × 15 grid, where the grid size is 1 mile × 1 mile.

(b) a sample synthetic concentration distribution corresponding to a single weather state (say as below, w_2) for which the wind direction was SW, wind speed 5 m/sec, stability was Pasquill category D, mixing height 1219m, and the ambient air temperature was taken as 284^{o}K.

(c) location of 3 hypothetical sampling stations x_1 = [84], x_2 = [127] and x_3 = [224].

(d) 10 other typical locations at which concentration estimates will be compared, viz x = [111], [113], [130], [143], [161], [164], [175], [178], [206], [209].

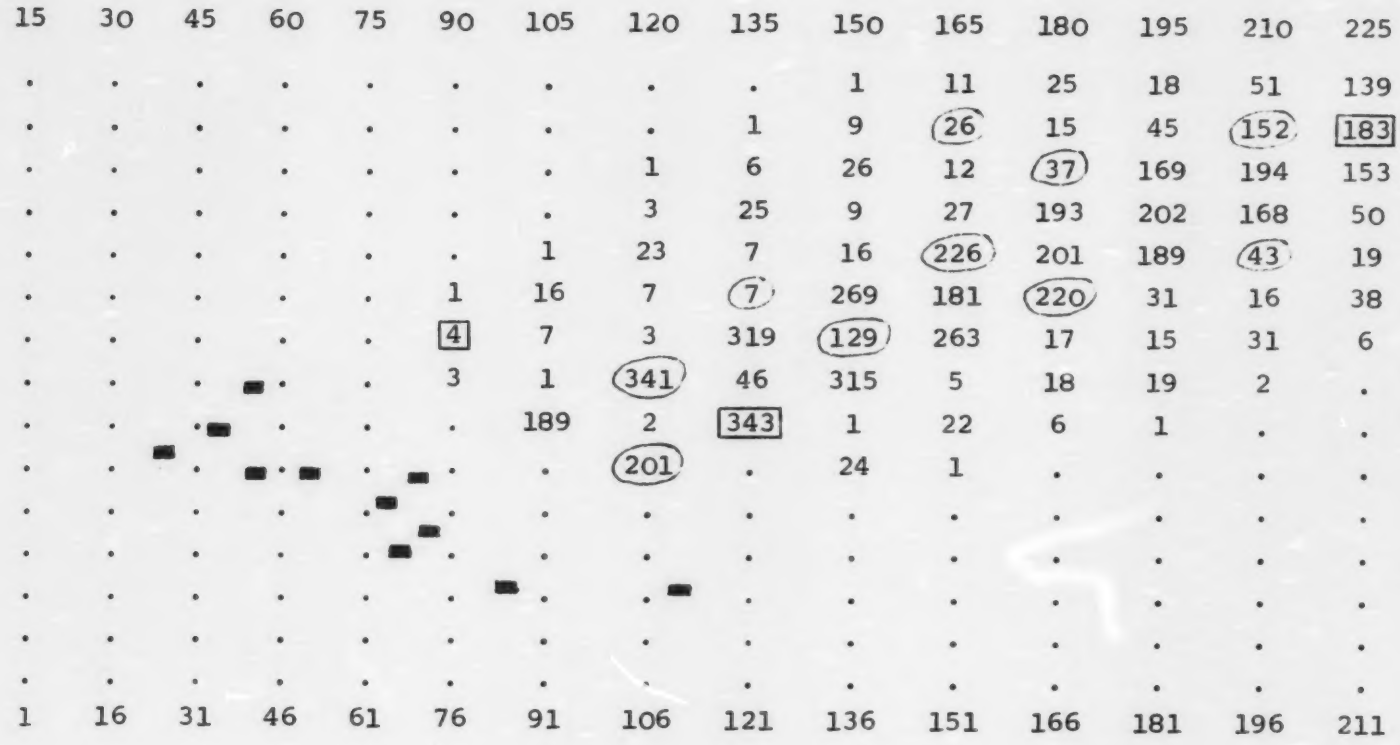
Synthetic concentration data were computed by the EPA code for all the weather states w in a hypothetical weather class

TABLE 1. HYPOTHETICAL SOURCE INVENTORY

SOURCES

NO	Q (G/SEC)	HP (M)	TS (DEG K)	VS (M/SEC)	D (M)	VF (M**3/SEC)	X (KM)	Y (KM)
1	191.1	61.0	600.0	6.1	2.6	32.2	9.190	6.300
2	47.7	63.6	600.0	4.8	2.9	31.9	9.190	6.300
3	21.1	30.5	811.0	29.2	0.9	18.6	9.190	6.300
4	14.2	38.1	727.0	9.2	1.7	20.3	9.190	6.300
5	7.0	38.1	727.0	7.0	2.1	25.9	9.190	6.300
6	59.2	21.9	616.0	4.3	2.0	12.9	9.190	6.300
7	87.2	61.0	616.0	5.2	2.1	18.7	9.190	6.300
8	25.3	36.6	477.0	11.9	2.7	66.6	8.520	7.840
9	101.0	36.6	477.0	16.0	2.0	49.3	8.520	7.840
10	41.6	18.0	727.0	9.0	2.6	47.5	8.520	7.840
11	222.7	35.7	477.0	5.7	2.4	26.6	8.050	7.680
12	20.1	45.7	727.0	2.4	1.9	6.9	8.050	7.680
13	20.1	50.3	727.0	1.6	1.5	2.8	8.050	7.680
14	20.1	35.1	727.0	1.5	1.6	2.9	8.050	7.680
15	20.0	34.7	727.0	1.6	1.5	2.8	8.050	7.680
16	24.7	30.0	727.0	9.0	2.2	34.2	9.190	6.300
17	67.5	76.3	473.0	10.7	3.0	78.2	5.770	10.810
18	66.7	82.0	603.0	12.9	4.4	196.1	5.620	9.820
19	63.7	113.0	546.0	9.3	5.2	196.7	4.600	9.500
20	6.3	31.0	460.0	5.0	1.6	10.1	8.230	8.870
21	36.2	50.0	460.0	7.0	2.2	26.6	8.750	5.880
22	28.8	50.0	460.0	7.0	2.5	34.4	11.240	4.560
23	8.4	31.0	460.0	5.0	1.6	10.1	6.140	8.780
24	172.4	42.6	616.0	13.4	4.6	219.8	14.330	6.200
25	171.3	42.6	616.0	16.1	3.7	169.4	14.330	6.200

Q : emission rate
HP : physical stack height
TS : stack gas temperature
VS : stack gas exit velocity
D : inside stack diameter
VF : stack gas volumetric flow rate
X : x-coordinate of stack
Y : y-coordinate of stack



▬ ≡ approximate location of individual sources

▭ ≡ sampling stations (3) x_1 = [84], x_2 = [127], x_3 = [224]

⬭ ≡ 10 additional typical points x = [111], [113], [130], [143], [161], [164], [175], [178], [206], [209]

Figure 1: Location of sources, sampling stations and 10 additional interpolation positions. Entries are concentration values ($\mu g/m^3$) estimated by EPA code, for one weather state w_2.

Ω_k defined to comprise the same wind direction, mixing height (and ambient air temperature) as for Figure 1, but with

wind speeds of 3, 5 and 7 m/sec

stability categories B, D, and E.

With these synthetic data, the 3 interpolation functions $m_i(x)$, $(i = 1,2,3)$, were then computed through least squares solution of the over-determined system of equations

$$c^w(x) = \sum_{i=1}^{3} m_i(x)\, c_i^w, \quad w \in \Omega_k,$$

where the "measurement" or "sampling" locations were as in Fig. 1(c), and the locations x were the 10 typical values selected in Fig. 1(d). The values of $m_i(x)$ are given in Table 2.

TABLE 2. COMPUTED INTERPOLATION FUNCTIONS $m_i(x)$ FOR THE WEATHER CLASS Ω_k AS DEFINED IN THE TEXT AND FOR THE 3 SELECTED SAMPLING LOCATIONS, AND 10 HYPOTHETICAL INTERPOLATION LOCATIONS

x	$m_1(x)$	$m_2(x)$	$m_3(x)$
[111]	3.8520	.8170	- .5307
[113]	.0637	.8866	.1718
[130]	.3159	.0119	.0040
[143]	- .4881	.5729	.6641
[161]	- .1281	-.0104	1.2720
[164]	1.2610	.0390	.0756
[175]	.3600	.0557	- .0303
[178]	.3355	.0837	.0265
[206]	.3563	.1097	.0033
[209]	.1367	-.1397	1.1000

As a numerical test of the suggested interpolation procedure, the computed interpolation function values were then used to estimate concentration values for certain selected weather states w in the weather class Ω_k, at the 10 selected locations, using the interpolation formula

$$\hat{C}^w(x) = \sum_{i=1}^{3} m_i(x) C_i^w$$

where C_i^w here denotes the "measured" values at the 3 sampling locations. The results are shown in Table 3 for two weather states w_1, w_2 differentiated as follows:

w_1 - wind speed 3 m/sec, stability category 5

w_2 - wind speed 5 m/sec, stability category 4,

but with all other weather conditions (wind direction, mixing height and air temperature being the same). In Table 3 the interpolated values $\hat{C}^w(x)$ are compared with the synthetic values $C^w(x)$ generated by the EPA code. For both weather states a very close agreement is obtained,* covering a wide range of concentration values. This is a very encouraging result and particularly so in relation to the rather complex structure of the typical concentration field (shown in Figure 1 for weather state w_2) and

---

*In an expanded numerical investigation the generally accepted statistical measures for the performance of the regressions would be reported, see for example Draper and Smith (1966). Such measures are among the output of any of the standard computer statistical packages, such as the IAREG Statistical Package, loc. cit.

the fact that the interpolation estimates only utilize 3 widely separated measurement locations. From examination of Figure 1, it is fairly obvious that in the absence of the interpolation scheme that has been developed and with only measurements at the 3 sampling stations, it would be virtually impossible to predict with any accuracy the fine structure of the concentration field.

TABLE 3. COMPARISON FOR THE SELECTED CONCENTRATION INTERPOLATION LOCATIONS, BETWEEN INTERPOLATED VALUES $\hat{c}^{w}(x)$ BASED ON ONLY 3 SAMPLING STATIONS AND THE SYNTHETIC VALUES $c^{w}(x)$ GENERATED BY THE EPA MODEL. COMPARISON IS FOR WEATHER STATES w_1 AND w_2, BOTH ASSUMED IN THE SAME WEATHER CLASS Ω_k IN UNITS $\mu gm/m^3$

x	$c^{w_1}(x)$	$\hat{c}^{w_1}(x)$	$c^{w_2}(x)$	$\hat{c}^{w_2}(x)$
[111]	63.78	54.08	200.77	196.09
[113]	339.20	356.15	340.69	335.43
[130]	4.24	5.64	7.00	5.84
[143]	447.37	451.59	314.47	315.71
[161]	508.65	506.53	225.27	227.61
[164]	44.32	43.02	25.63	27.55
[175]	3.86	6.10	16.74	14.76
[178]	33.59	37.93	36.85	34.62
[206]	32.25	37.06	42.11	39.38
[209]	401.70	395.83	151.40	152.83

SECTION 8

THE SAME EXAMPLE WITH AGGREGATION OF PSEUDO-SOURCES

As a simple example the 25 "real" sources of Table 1 in the previous section were subdivided rather crudely into only 4 classes based on the magnitudes of the individual strengths, as shown in Table 4. A more detailed analysis would produce more pseudo-source classes with smaller within-class variation.

TABLE 4. SELECTION OF PSEUDO-SOURCE CLASSES BY LEVEL OF EMISSIONS

Pseudo-Source	Sources in Class	Range of Individual Source Strengths in Class	Number of Sources
1	4,5,20,23	6- 14	4
2	3,8,12-16,21,22	20- 36	9
3	2,6,7,10,17-19	41- 68	7
4	1,9,11,24,25	101-222	5

For the experiment the aggregated-averaged approximation was employed, corresponding to the "pure" dispersion function

$$u_r^w(x) = \sum_{r\ \text{class}} v_j^w(x)/\text{No. in class};\ r = 1,2,3,4.$$

These were computed for $r = 1,2,3,4$ and each of the 9 weather

states w in the weather class Ω_k defined in Section 7. As before, the EPA model was used to translate the source-positional and meteorological information into values of $u_r^w(x)$, for the same three sampling locations x_1, x_2 and x_3 and the same ten receptor locations as in Figure 1. However, in the present case the system of equations (26) (in contrast to the system (27) for Section 7) were solved by least-squares for the three interpolant functions $m_i(x)$, $i = 1,2,3$. These values are given in Table 5.

TABLE 5. COMPUTED INTERPOLATION FUNCTIONS $m_i(x)$ FOR THE WEATHER CLASS Ω_k, AND FOR THE 3 SELECTED SAMPLING LOCATIONS, 4 SELECTED PSEUDO-SOURCE CLASSES, AND 10 HYPOTHETICAL INTERPOLATION LOCATIONS

x	$m_1(x)$	$m_2(x)$	$m_3(x)$
[111]	1.7800	.6172	- .3040
[113]	.1086	-.4374	1.9360
[130]	- .1512	.2402	- .1580
[143]	- .3752	.9025	.2945
[161]	- .02555	-.3024	1.7920
[164]	.1736	.1080	- .05258
[175]	.1321	.0340	.003239
[178]	- .1658	.3158	- .07981
[206]	.08797	.08334	.03493
[209]	.0311	-.1820	1.2740

Then, as in Section 7 (see Table 3), the computed interpolation functions of Table 5 were used to estimate the concentration values for certain selected weather states in the weather class Ω_k at the 10 selected locations using the formula

$$\hat{c}^{w}(x) = \sum_{i=1}^{3} m_i(x)\, c_i^{w},$$

where c_i^{w} here denotes the "measured" values at the 3 sampling locations. Analogous to Table 3, the interpolated values $\hat{c}^{w}(x)$ are compared with the synthetic values $c^{w}(x)$ generated by the EPA model in Table 6.

TABLE 6. COMPARISON FOR THE SELECTED CONCENTRATION INTERPOLATION LOCATIONS, BETWEEN INTERPOLATED VALUES $\hat{c}^{w}(x)$ (BASED ON ONLY 3 SAMPLING STATIONS AND 4 PSEUDO-SOURCES) AND THE SYNTHETIC VALUES $c^{w}(x)$ GENERATED BY THE EPA MODEL. COMPARISON IS FOR WEATHER STATES w_1 AND w_2 BELONGING TO WEATHER CLASS Ω_k (WITH UNITS $\mu gm/m^3$).

x	$c^{w_1}(x)$	$\hat{c}^{w_1}(x)$	$c^{w_2}(x)$	$\hat{c}^{w_2}(x)$
[111]	63.78	79.08	200.77	162.08
[113]	339.20	634.53	340.69	202.89
[130]	4.24	14.39	7.00	53.08
[143]	447.37	410.25	314.47	361.78
[161]	508.65	620.46	225.27	222.51
[164]	44.32	14.01	25.63	28.02
[175]	3.86	9.79	16.74	11.50
[178]	33.59	70.22	36.85	93.14
[206]	32.35	41.05	42.11	35.22
[209]	401.70	451.82	151.40	169.67

Some of the comparisons of Table 6, that is based on the aggregated-averaged approximation, are good and some are poor, although overall the agreement may be considered very encouraging in view of the crude initial aggregation of the sources into 4 classes with unduly large variations of source strengths within

each class. It is perhaps even surprising that without any knowledge of the real source-strengths (since for the aggregated-averaged approximation the functions $u_r^w(x)$ are "pure" dispersion functions), except a very crude grouping by general level of emission, the numerical features of the concentration field are so well produced by the interpolation procedure. Of course, the errors arising in the aggregation approximation that was used can be avoided if an unaggregated pseudo-source distribution is employed. As we have seen, this will involve more computation to produce the required interpolation functions for the real world, where the number of sources may be quite large, although still of a magnitude well within the capabilities of modern computers. In any case the interpolation functions may be precalculated for subsequent use in a real-time air quality modeling situation.

SECTION 9

RECOMMENDATIONS

The results reported herein relate to a small-scale and limited exploratory study. They point to the possibility for innovative and optimal interpolation procedures for real-time estimation of the fine structure of the air quality distribution over a region, in terms of sparse measurement data. The first results are very encouraging and are suggestive of considerable operational value in air quality management. Further effort should be directed towards analysis of several theoretical questions that it was not possible to resolve completely in a limited study. These include an analysis of the degree of independence of the interpolation functions from the spatial distribution of pollutant source-strengths (in contrast to purely source-positional parameters) together with an analysis of the errors of interpolation and the development of optimal operational formulae. At the same time the techniques developed should be tested against real air pollution data that is becoming available from appropriate field programs, such as the EPA Regional Air Pollution Study.

REFERENCES

Brier, G. W., 1973. Validity of the Air Quality Display Model Calibration Procedure, Environmental Protection Agency, EPA-R4-73-017.

Calder, K. L., et.al., 1975. Empirical Techniques for Analyzing Air Quality and Meteorological Data: Part II. Feasibility of a Source-Oriented Empirical Air Quality Simulation Model, Environmental Protection Agency, ESRL-RTP-054, Series No. 4.

Calder, K. L., 1976. Multiple-Source Plume Models for Urban Air Pollution - Their General Structure, NATO/CCMS Panel on Modeling, Documentation for Practical Demonstration of Urban Air Quality Simulation Models.

Dahlquist, G. and Björck, Å., 1974. Numerical Methods (Transl. by N. Anderson), Prentice-Hall, Englewood Cliffs, N. J.

Draper, N. R. and Smith, H., 1966. Applied Regression Analysis, J. Wiley and Sons, New York.

Heimback, J. A. and Sasaki, Y., 1975. A Variational Technique for Mesoscale Objective Analysis of Air Pollution, J. Appl. Meteorology 14, 194-203.

Instructions for Using the IA Regression Package, 1974. Graduate School of Industrial Administration, Carnegie-Mellon University, Pittsburgh, Pa. 15213.

Hrenko, J. M. and Turner, D. B., 1975. An Efficient Gaussian-Plume Multiple-Source Air Quality Algorithm, #75-04.3, 68th Annual Meeting of APCA, Boston, Massachusetts, June 15-20.

Lanczos, C., 1965. Applied Analysis, Prentice-Hall, Englewood Cliffs, New Jersey.